

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligentech, Inc.

Sequence: A. Comp: 1 CGCCAGGCCGCTCAA..... TCAACTGGAAAAAAA 152  
  1527 GCGGTGCGGGCGGAGTT..... AATGTGACCTTTTTTT

batch STD : Dbase 0; Query 0

post-processing: Minimum Match 0% Listing first 45 summaries

database:  
enews  
1:BCT 2:FUN 3:GEN1 4:GEN2 5:HTG1 6:HTG2 7:HUM 8:INV  
9:ORG 10:MAM 11:VRT 12:PLN 13:PRO 14:ROD 15:SYN 16:UNC  
17:VIR

18:BCT1 19:BCT2 20:BCT3 21:BCT4 22:BCT5 23:BCT6 24:BCT7  
25:BCT8 26:BCT9 27:BCT10 28:BCT11 29:BCT12 30:GEN1  
31:GEN3 32:HNG1 33:HTG2 34:HNG3 35:INV1 36:INV2 37:INV3  
38:INV4 39:INV5 40:INV6 41:INV7 42:INV8 43:INV9 44:INV10

45:INVL 46:IMAM1 47:MAM2 48:IMAM3 49:VRT1 50:VRT2 51:VRT3  
52:VRT4 53:PAT1 54:PAT2 55:PAT3 56:PAT4 57:PAT5 58:PHG  
59:PLN1 60:PLN2 61:PLN3 62:PLN4 63:PLN5 64:PLN6 65:PLN7  
66:PLN8 67:PLN9 68:PLN10 69:PLN11 70:PR1 71:PR12

79;PR110 80;PR11 81;PR12 82;PR13 83;PR14 84;PR15  
 85;ROD1 86;ROD2 87;ROD3 88;ROD4 89;ROD5 90;ROD6 91;ROD7  
 92;ROD8 93;STR 94;SYN 95;UMA 96;VRLL9 97;VRL2 98;VRL3  
 99;VRL4 100;VRL5 101;VRL6 102;VRL7 103;VRL8 104;VRL9

base: database: genbank-new3  
106:BCT 107:GEN1 108:GEN2 109:HTG1 110:HTG2 111:INV  
112:MAR 113:MRT 114:PHG 115:PLN 116:PRIL 117:PR12  
118:ROD 119:SYN 120:TTNA 121:VRT

statistics: 122:part1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
C	1	32	2.1	215	57	I28278 Sequence 5 from patent 1-31e-04	4.58e-05
C	2	30	2.0	215	57	I28278 Sequence 5 from patent 1-33e-04	3.33e-04
C	3	28	1.8	354	111	OFU89259 Oxytricha fallax 57kd	3.33e-03
C	4	28	1.8	354	8	OFU89259 Oxytricha fallax 57kd	3.33e-03
C	5	25	1.6	354	111	OFU89259 Oxytricha fallax 57kd	3.26e-03
C	6	25	1.6	354	8	OFU89259 Oxytricha fallax 57kd	3.26e-03
C	7	25	1.6	1441	51	S66477 glial fibrillary acid	3.26e-01
C	8	25	1.6	1933	49	CRAGFL Carassius auratus (GF	3.26e-01
C	9	24	1.6	7095	92	RMRP5 R. norvegicus mRNA for P. sativum gene for ch	1.38e+01
C	10	24	1.6	7322	64	PSCHSAB PSCHSAB	1.38e+00
C	11	24	1.6	18632	35	CEC29FB Caenorhabditis elegan	1.38e+00
C	12	24	1.6	47745	8	CET09DA H.sapiens LIPA gene,	1.38e+00
C	13	24	1.6	115419	34	HS799FF.0 Human DNA sequence	1.38e+00
C	14	24	1.6	150894	20	D90904 Synechocystis sp. PCC	1.38e+00
C	15	23	1.5	1344	92	ECACEA E. coli aceA gene cod	5.59e+00
C	16	23	1.5	1747	21	ECACERA Escherichia coli 1soc	5.59e+00
C	17	23	1.5	1851	117	HSLIP4 H.sapiens LIPA gene,	5.59e+00
C	18	23	1.5	1851	57	HSLIP4 H.sapiens LIPA gene,	5.59e+00
C	19	23	1.5	2358	21	ECOIDKPA E.Coli ackA gene enco	5.59e+00
C	20	23	1.5	3021	69	YSCGLM3 S.cerevisiae nitrogen	5.59e+00
C	21	23	1.5	3155	122	SCEN973 DNA expression cassette	5.59e+00
C	22	23	1.5	3177	43	LISRNARP L.stagnalis mRNA for	5.59e+00
C	23	23	1.5	3720	20	ECBQW9 E. coli chromosomal r	5.59e+00
C	24	23	1.5	6000	122	ECBQW9 E. coli chromosomal r	5.59e+00
C	25	23	1.5	14833	21	ECAE000474 Escherichia coli from	5.59e+00
C	26	23	1.5	38586	37	CFEL55C12 Caenorhabditis elegan	5.59e+00
C	27	23	1.5	43100	65	SCEN973 Saccharomyces cerevisiae	5.59e+00
C	28	23	1.5	145750	33	HS11JQ24 Human DNA sequence	5.59e+00
C	29	23	1.5	176195	42	ECBQW9 E. coli chromosomal r	5.59e+00
C	30	23	1.5	618	108	FRO002420 F.rubripes GSS sequen	2.15e+01
C	31	22	1.4	618	3	FRO002420 F.rubripes GSS sequen	2.15e+01
C	32	22	1.4	1230	67	SLU39443 S.cerevisiae chromosome	2.15e+01
C	33	22	1.4	1245	61	CREPFRSI C.reinhardtii esculent	2.15e+01
C	34	22	1.4	1256	47	OCSPIGN O.cuniculus 5P'7 gene	2.15e+01
C	35	22	1.4	1761	67	SCYO085C S.cerevisiae chromosome	2.15e+01
C	36	22	1.4	2492	65	SCDNATRS2 S.cerevisiae fus2 gene	2.15e+01
C	37	22	1.4	3699	67	SCIOL084W S.cerevisiae chromosome	2.15e+01
C	38	22	1.4	3796	90	MUSTNTR1 Murine tumor necrosis	2.15e+01
C	39	22	1.4	17013	40	EAMMSL E.amylovora (E77/74)	2.15e+01
C	40	22	1.4	40267	77	HSU19E9 Human DNA sequence	2.15e+01
C	41	22	1.4	90468	34	HS465Q10 Human DNA sequence	2.15e+01
C	42	22	1.4	99370	15	ATAC00103 Genomic sequence for	2.15e+01
C	43	22	1.4	99370	5	ATAC003 Genomic sequence for	2.15e+01
C	44	22	1.4	99370	32	ATAC00103 *** SEQUENCING IN PRO	2.15e+01
C	45	22	1.4	257912	103	ATAC00103 Paramaecium bursaria C	2.15e+01
ALIGNMENTS							
FEATURES	source	Location/Qualifiers					
REFERENCE		1. (bases 1 to 215)					
AUTHORS	Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.						
TITLE	Plant inhibitors of fungal polygalacturonases and their use to control fungal disease						
JOURNAL	Patent: US 5569830-A 5 22-OCT-1996;						
BASE COUNT	ORIGIN	1. 215 /organism="unknown" 8 C 25 g	26 t	141 others			

Query Match	2.1%; Score 32; DB 57; Length 215;
Best Local Similarity	18.8%; Pred. No. 4.58e-06;
Matches	26; Conservative 54; Mismatches 56; Indels 2; Gaps 2;
JOURNAL	JOURNAL 94134747
MEDLINE	MEDLINE 2
REFERENCE	REFERENCE (bases 1 to 354)
AUTHORS	Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.
TITLE	TITLE Selection on the protein-coding genes of the TBEL family of transposable elements in the ciliates Oxytricha fallax and O. trifallax
JOURNAL	JOURNAL Unpublished
REFERENCE	REFERENCE 3 (bases 1 to 354)
AUTHORS	AUTHORS Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.
TITLE	TITLE Direct Submission
JOURNAL	JOURNAL Submitted (11-FEB-1997) Oncological Science, University of Utah, School of Med. RmC334, USA, UT 84132, USA
FEATURES	FEATURES Location/Qualifiers
SOURCE	SOURCE 1..354
FEATURES	/organism="Oxytricha fallax"
SOURCE	/strain="9D1"
FEATURES	/transposon="TBEL"
SOURCE	/note="this is a bulk sequence that was generated from a PCR product that represents many transposon templates"
BASE COUNT	BASE COUNT 1..354
CDS	CDS <1..>354
FEATURES	/codon_start=1
SOURCE	/product="57KD zinc finger/protein chimera"
FEATURES	/ab_xref="PID:gi1881676"
SOURCE	/translation="HTDLXKHILAKKKKXXXXXXKLXKRRAREXXXXXXQAXEOKXXKRNKXKSEXTKIMIKIYKIPVLAQIDLTSQSYLILEDSDFKVIKHDHOAYKVKPFNLY"
BASE COUNT	BASE COUNT 106 a 42 c 41 g 54 t 111 others
ORIGIN	ORIGIN
RESULT	Query Match 1.8%; Score 28; DB 111; Length 354;
ID	ID Best Local Similarity 30.1%; Pred. No. 3.33e-03;
AC	AC Matches 25; Conservative 33; Mismatches 24; Indels 1; Gaps 1;
NT	NT 91881675
DT	DT 13-MAR-1997 (Rel. 51, Created)
DR	DR 13-MAR-1997 (Rel. 51, Last updated, Version 1)
DE	DE Oxytricha fallax 57kd zinc finger/protein chimera gene, partial cds.
KW	KW
OC	OC Oxytricha fallax
OC	OC Eukaryotes; mitochondrial eukaryotes; Alveolata; Ciliophora; hypotrichids; Stichotrichida; Oxytrichida; Oxytricha.
RN	RN [1]
RP	RP 1..354
RX	RX MEDLINE; 94134747.
RA	RA Doak,T.G., Doerder,F.P., Jahn,C.L., Herrick,G.;
RT	RT "A proposed superfamily of transposase genes: transposon-like elements in ciliated Protozoa and a common 'D35E' motif;"
RL	RL Proc. Natl. Acad. Sci. U.S.A. 91:942-945(1994).
RN	RN [2]
RP	RP 1..354
RA	RA Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J., Herrick,G.;
RT	RT "Selection on the protein-coding genes of the TBEL family of transposable elements in the ciliates Oxytricha fallax and O. trifallax";
RT	RT transposable elements in the ciliates Oxytricha fallax and O.
RL	RL Unpublished.
RN	RN [3]
RP	RP 1..354

	CDS
RA	Doak T.G., Williams K., Witherspoon D.J., Herrick G.;
RT	; Submitted (11-FEB-1997) to the EMBL/GenBank/DDBJ databases.
RL	Oncological Science, University of Utah, School of Med. Rm5C334, USA, UT 84132, USA
RL	Location/Qualifiers
FH	
FT	source
FT	1..354
FT	/organism="Oxytricha fallax"
FT	/strain="9DL"
FT	/transposon="TBEI"
FT	/note="this is a bulk sequence that was generated from a PCR product that represents many transposon templates"
FT	<1..>354
FT	/codon_start=1
FT	/product="57kD zinc finger/protein chimera"
FT	/transl_table=6
FT	/db_xref="PID:9181676"
D	/translation="HTRDIXKHLKAHKXXEXXXXXKLXKRAREXXXXXX
D	AXXXXVXXRNXLXSEXTKIMKIQVKPVLAQIDLTSLQSYLIEDSFDDKVIX
D	HQYAVKVPFNYIK"
FT	Sequence 354 BP; 106 A; 42 C; 41 G; 54 T; 111 other;
FT	Query Match 1..8%; Score 28; DB 8; Length 354;
FT	Best Local Similarity 30.1%; Pred. No. 3.33e-03;
FT	Matches 25; Conservative 33; Mismatches 36; Indels 0; Gaps 0;
FT	Accession 089259
FT	Version 24; Indels 1; Gaps 1;
FT	Keywords
FT	Organism Oxytricha fallax.
FT	Eukaryota; Mitochondrial eukaryotes; Alveolata; Ciliophora; Hypotrichia; Stichotrichida; Oxytrichida; Oxytricha.
FT	RN [1]
FT	RESULT 5
LOCUS	ORF089259 354 bp DNA_finger/protein chimera gene, partial
DEFINITION	INV
ACCESSION	089259
NID	9181675
KEYWORDS	.
SOURCE	Oxytricha fallax.
ORGANISM	Oxytricha fallax.
REFERENCE	1 (bases 1 to 354)
AUTHORS	Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G.
TITLE	A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common 'D35E' motif
JOURNAL	PROC. NATL. ACAD. SCI. U.S.A. 91 (3), 942-945 (1994)
MEDLINE	9413477
REFERENCE	2 (bases 1 to 354)
AUTHORS	Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.
TITLE	Selection on the protein-coding genes of the TBEI family of transposable elements in the ciliates Oxytricha fallax and O. trifallax
JOURNAL	UNPUBLISHED
REFERENCE	3 (bases 1 to 354)
AUTHORS	Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.
TITLE	Direct Submission
JOURNAL	Submitted (11-FEB-1997) Oncological Science, University of Utah, School of Med. Rm5C334, USA, UT 84132, USA
FEATURES	Location/Qualifiers
FT	source
FT	1..354
FT	/organism="Oxytricha fallax"
FT	/strain="9DL"
FT	/transposon="TBEI"
FT	/note="this is a bulk sequence that was generated from a PCR product that represents many transposon templates"
FT	<1..>354
FT	/codon_start=1
FT	/product="57kD zinc finger/protein chimera"
FT	/transl_table=6
FT	/db_xref="PID:9181676"
FT	/translation="HTRDIXKHLKAHKXXEXXXXXKLXKRAREXXXXXX
FT	AXXXXVXXRNXLXSEXTKIMKIQVKPVLAQIDLTSLQSYLIEDSFDDKVIX
FT	HQYAVKVPFNYIK"
FT	PCR product that represents many transposon templates"

D  
 FT  
 SQ      Sequence 354 BP; 106 A; 42 C; 41 G; 54 T; 111 other;  
 Query Match 1.6%; Score 25; DB 8; Length 354;  
 Best Local Similarity 23.7%; Pred. No. 3.26e-01;  
 Matches 23; Conservative 38; Mismatches 36; Indels 0; Gaps 0;  
 Db 99 rgcdagrrgargyhyhargwbgwydyggcyykayaaagcwagmgarwasyaswtrygaar 158  
 :|||::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Cp 799 GGAAGCTGTGTTGGAGCTGTTGCCGTTTACCTCTACTCTTGCCGTTA 740  
 :|||:|||:|||:|||:|||:|||:|||:|||:  
 Cp 739 CTTTGGCCAGCAATTGCTACACAGACATAAG 703  
 :|||:  
 RESULT 7      HOYAKVYPENYLX"  
 LOCUS S66477 1441 bp mRNA VRT 17-DEC-1993  
 INITIATION glial fibrillary acidic protein [clone 2] [Cyprinus carpio=carp,  
 ORGANISM Cyprinus carpio]  
 AUTHORS Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Actinopterygii; Neopterygii; Teleostei; Buteleosteii;  
 Ostariophysi; Cypriniformes; Cyprinoidea; Cyprinidae; Cyprininae;  
 TISSUE brain, mRNA Partial, 1441 nt].  
 ACCESSION S66477  
 NID 9435738  
 KEYWORDS common carp brain.  
 SOURCE  
 REFERENCE 1 (bases 1 to 1441)  
 AUTHORS Cohen,I.; Shani,Y. and Schwartz,M.  
 TITLE Cloning and characteristics of fish glial fibrillary acidic  
 protein: implications for optic nerve regeneration  
 JOURNAL J. Comp. Neurol. 334 (3), 431-443. (1993)  
 MEDLINE 93388923  
 REMARK GenBank staff at the National Library of Medicine created this  
 entry [NBL1 gibbsq 1390569] from the original journal article.  
 This sequence comes from Fig. 1.  
 FEATURES Location/Qualifiers  
 Source  
 CDS  
 REFERENCE 1.1.141  
 /organism="Cyprinus carpio"  
 1..635  
 /part=1  
 /note="Description: glial fibrillary acidic protein, GFAP;  
 Method: conceptual translation with partial peptide  
 sequencing. This sequence comes from Fig. 2."  
 /codon\_start=1  
 /product="glial fibrillary acidic protein"  
 /db\_xref="PID:435739"  
 /translation="VDLVERPDILTALKERIAFPEAMATSNMOTTEEWRSKFADLT  
 DAASRTTEALRQAKENNRQIGCITCDESLSGSNEELERQRMEERFTETAG  
 YDTTARLELDIOMKREEMAHQLQDOLINKVLAIDETATYRKILEGESRITVPPV  
 QNFNTIQFRDSLKDILTPRHVKYSIVVREYETDGEELIKEESTERNIDP"  
 BASE COUNT 420 a 273 c 347 g 401 t  
 ORIGIN  
 Query Match 1.6%; Score 25; DB 51; Length 1441;  
 Best Local Similarity 77.8%; Pred. No. 3.66e-01; Mismatches 10; Indels 0; Gaps 0;  
 Matches 35; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 Db 438 gaactgtcgaaaggagagaatctgtccgtgtca 482  
 :|||:|||:|||:|||:|||:  
 Cp 931 GAAGCCTCTCGAGGAGGTAGCTCCATGCCGCTGCA 887  
 :|||:  
 RESULT 8      CRGP1 1933 bp mRNA VRT 01-SEP-1993  
 LOCUS Carassius auratus (GFAP-1) mRNA, complete cds.  
 DEFINITION L23076  
 ACCESSION 9388622  
 NID  
 KEYWORDS  
 SOURCE Carassius auratus adult retina cDNA to mRNA.

ORGANISM Carassius auratus  
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Actinopterygii; Neopterygii; Teleostei; Buteleosteii;  
 Ostariophysi; Cypriniformes; Cyprinoidea; Cyprinidae; Cyprininae;  
 Carassius, 1 (bases 1 to 1933);  
 Glasgow,E. and Schechter,N.  
 REFERENCE 1  
 AUTHORS  
 TITLE Nucleotide sequence of a GFAP - like intermediate filament cDNA  
 from Goldfish retina  
 JOURNAL Unpublished (1993)  
 FEATURES Location/Qualifiers  
 source 1.1.133  
 /Organism="Carassius auratus"  
 /dev\_stage="adult"  
 /tissue\_type="retina"  
 20..1099  
 CDS  
 REFERENCE 1.1.133  
 /gene="GFAP-1"  
 /note="putative"  
 /codon\_start=1  
 /db\_xref="PID:9388623"  
 /translation="MGLNDRFASYIEKVRFLFQNLKMLVAELNLRGKEPSRGDIQ  
 BELRSIRROVGLNAGKARIEELDRNLDQDATERKORQENALRQEENLNMPFDQ  
 VDEALANRVLQELKIDOLEEISLTKVHEEEMOLOEOLAOQHVOHLDVSUPDLT  
 ALKERAQFAMASNMOTFEMWRSRADLTDAGRNALRAKQANERAKRQ  
 DDCDLESRSRNEESIERSERFALENPGYDQTFVABLEDELOMKBEMARHILQY  
 ODLLNKVLAIDETATYRKILEGESRITVPPVQNTNQFRDTSLDTKLTPAEHKRS  
 IIVWRVETRQEBIKESTERKDLP"  
 BASE COUNT 561 a 395 c 507 g 470 t  
 ORIGIN  
 REFERENCE 1.1.133  
 /note="putative"  
 /codon\_start=1  
 /db\_xref="PID:9388623"  
 /translation="MGLNDRFASYIEKVRFLFQNLKMLVAELNLRGKEPSRGDIQ  
 BELRSIRROVGLNAGKARIEELDRNLDQDATERKORQENALRQEENLNMPFDQ  
 VDEALANRVLQELKIDOLEEISLTKVHEEEMOLOEOLAOQHVOHLDVSUPDLT  
 ALKERAQFAMASNMOTFEMWRSRADLTDAGRNALRAKQANERAKRQ  
 DDCDLESRSRNEESIERSERFALENPGYDQTFVABLEDELOMKBEMARHILQY  
 ODLLNKVLAIDETATYRKILEGESRITVPPVQNTNQFRDTSLDTKLTPAEHKRS  
 IIVWRVETRQEBIKESTERKDLP"  
 RESULT 9      RRMAB15 7095 bp RNA ROD 21-OCT-1992  
 LOCUS R-norvegicus mRNA for microtubule associated protein IB.  
 DEFINITION R-norvegicus mRNA for microtubule associated protein IB.  
 ACCESSION X60370 X60371 X60550  
 NID 957018  
 KEYWORDS MAPB gene; microtubule-associated protein.  
 SOURCE  
 ORGANISM Rattus norvegicus  
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;  
 Murinae; Rattus.  
 REFERENCE 1  
 AUTHORS Launer,W., Kratz,J., Staunton,J., Feilck,P. and Wiche,G.  
 TITLE Identification of two distinct microtubule binding domains on  
 recombinant rat MAP 1B  
 JOURNAL Eur. J. Cell Biol. 57 (1), 66-74 (1992)  
 MEDLINE 92347374  
 REFERENCE 2  
 AUTHORS Wiche,G.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-AUG-1991). G. Wiche, Inst. of Biochemistry, University  
 of Vienna, Wehringerstrasse 17, 1090 Vienna, AUSTRIA  
 REFERENCE 3  
 AUTHORS Wiche,G.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-AUG-1992). G. Wiche, Institute of Biochemistry and  
 Molecular Biology, University of Vienna, Dr. Bohrgasse 9, 1030  
 Vienna, AUSTRIA  
 FEATURES Location/Qualifiers  
 source 1..7095  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /dev\_stage="adult"





REFERENCE  
1 (bases 1 to 115419)  
AUTHORS Hunt,A.  
TITLE Direct Submission  
JOURNAL Submitted (05-Nov-1996) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail: enquiries:  
hungary@sanger.ac.uk Clone requests: cloneresquestsanger.ac.uk  
COMMENT IMPORANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished sequence: BK799F10 Contig-ID: 00052 Length: 992 bp Unfinished sequence: BK799F10 Contig-ID: 00061 Length: 885 bp Unfinished sequence: BK799F10 Contig-ID: 00065 Length: 12614 bp Unfinished sequence: BK799F10 Contig-ID: 000949 Length: 830 bp Unfinished sequence: BK799F10 Contig-ID: 000953 Length: 785 bp Unfinished sequence: BK799F10 Contig-ID: 000960 Length: 958 bp Unfinished sequence: BK799F10 Contig-ID: 000968 Length: 990 bp Unfinished sequence: BK799F10 Contig-ID: 01506 Length: 6932 bp Unfinished sequence: BK799F10 Contig-ID: 01718 Length: 38510 bp Unfinished sequence: BK799F10 Contig-ID: 02165 Length: 13742 bp Unfinished sequence: BK799F10 Contig-ID: 02195 Length: 2254 bp Unfinished sequence: BK799F10 Contig-ID: 02223 Length: 1000 bp Unfinished sequence: BK799F10 Contig-ID: 02224 Length: 1000 bp Unfinished sequence: BK799F10 Contig-ID: 02225 Length: 1341 bp Unfinished sequence: BK799F10 Contig-ID: 02226 Length: 1016 bp Unfinished sequence: BK799F10 Contig-ID: 02214 Length: 19570 bp. \*\*\*  
\*\*\* WARNING: Phase 1 High Throughout Genome Sequence \*\*\*  
\* This sequence is unfinished. When sequencing is complete,  
\* by a single finished sequence with the same accession number.  
FEATURES source  
BASE COUNT 24169 a 28425 c 26764 g 23990 t 12071 others  
ORIGIN

Query Match 1.6%; Score 24; DB 34; Length 115419;  
st Local Similarity 80.0%; Pred. No. 1.38e-00; Mismatches 8; Indels 0; Gaps 0;  
ataches 32; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 109739 tctttgttttgtttttccacttcggatcgcaag 109778  
||| ||||| ||||| ||||| ||||| ||||| ||||| 794

CP 83.3 TCTTTGTGTTGATCTTCATACTCTACAGCAGCGAAG

RESULT 14 D90904 150894 bp DNA BCT 20-NOV-1996  
LOCUS Synechocystis sp. PCC6803 complete genome, 6/27, 530555-781448.  
DEFINITION  
ACCESSION D90904  
NID 6152225  
KEYWORDS formaldehyde dehydrogenase; ABC transporter; DNA ligase; DNA polymerase III alpha subunit; GmB protein; LYSR transcriptional regulator; Mg-protoporphyrin IX; N-Acetylornithine aminotransferase; NADH dehydrogenase subunit 4; NADH dehydrogenase subunit 5; acetate kinase; aspartoylase(ASP); cation-transporting ATPase; cell division protein FtsH; chemotaxis protein CheA; cytochrome oxidase subunit I; cytochrome oxidase d subunit II; dienelactone hydrolase; dihydroflavonol 4-reductase; dihydropteroate pyrophosphorylase; esterase; ferrous iron transport protein B; fibrillin; formaldehyde dehydrogenase (glutathione); high light-inducible protein; leader peptidase I; lysostaphin; malonyl coenzyme A acyl carrier protein transacylase; methyl accepting chemotaxis protein (MCP); oligopeptide transport ATP-binding

SOURCE  
ORGANISM Synechocystis sp. (strain:PCC6803) DNA.  
Eubacteria; Cyanobacteria; Chroococcales; Synechocystis.  
REFERENCE 1 (bases 1 to 150894)  
AUTHORS Tabata,S.  
TITLE JOURNAL Submitted (28-JUN-1996) to the DDBJ/EMBL/GenBank databases. Satoshi Tabata, Kazusa DNA Research Institute, Laboratory of Gene Structure (E-mail:tabata@kazusa.or.jp), Tel:+81-438-52-3933, Fax:+81-438-52-3934)  
COMMENT 2 (sites)  
Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirose,M., Sugiyama,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsunaga,A., Muraki,A., Nakaraki,N., Naruo,K., Okumura,S., Shimpou,S., Takeuchi,C., Wada,T., Watanabe,A., Yamada,M., Yasuda,M. and Tabata,S.  
TITLE Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions DNA Res. 3 (3), 109-136 (1996)  
97061201 Potential protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis.  
FEATURES source  
CDS  
CDSSource  
/product="ORF\_ID:s110998"  
/note="ORF\_ID:s110998"  
/codon\_start=1  
/product="LYR transcriptional regulator"  
/note="ORF\_ID:s111911"  
/product="DB\_XREF\_PDB:g1052226"  
/transl\_table=11  
/translation="MSDTPFTDQLRILKAIASEGSKRAADTLVSPAVSLQVNL  
EKOLSPVPLRDRGERKAQITAEGLNLNGNKEKITLCOMTCRATEDLQLQGGILIVGA  
SQTGTYLPRLPMIGMRCQYPOVYLOVHSIRRTANGVANSGYDAIIGGEVPAELO  
ETIVVLPDAEDEALYLVLPLAQAEIQKLFISLSQSOSTIRVKVILSQ  
GEIDTKRKTEMLNSTAKNAVQKISAFIREVLPQFSTHDALDPERLFAPIYSSNGDROGDGKD  
RVSVLTNPYRSKSAAFIREVLPQFSTHDALDPERLFAPIYSSNGDROGDGKD  
GKSIEIESTVT" complement(1342..1539)  
/note="ORF\_ID:ss11911"  
/codon\_start=1  
/product="hypothetical protein"  
/db\_xref="PID:g1052227"  
/transl\_table=11  
/translation="MNSTQQARALMMRHQFIKNRQMSLRAAEIGVEAEDFWTT  
VCKPQPSFRTRYDRSNASL" CDS  
/note="ORF\_ID:s1r1019"  
/product="hypothetical protein"  
/db\_xref="PID:g1052228"  
/transl\_table=11  
/translation="MFNSPVRPMYGFVTDQIPLRGGNPJAVPFDQEGTDFQMQ  
KIAEAINSETVFLPVPGPTTGNRLRIFTKBDLFGHTPTGTYLQLQPPSP  
VTTTWOLPEPGVGVPUVTHYKOCOLVCPTEVQLPOTPKSDSCDALLLGLSIDO  
LQGGEIPEQASLGFLPFLTPEELNEANRISTNSPWNQNLLAGQWADCVCYCLAPGDP  
SLSGQDNKUHGRMAPGIGTASPDATSGVAAVGLYQUDLDTPGSHWQIOBGKAL  
GRSQLOLTVKQGIGRAVKGVAGRSVLSSEGPNLGN"  
/product="tRNA-Leu"

tRNA

**CDS**  
*complement(3'235..4'326)*  
*/note="ORF\_ID:s110997"*  
*/codon\_start=1*  
*/product="hypothetical protein"*  
*/db\_xref="PID:91652229"*  
*/transl\_table=11*  
*/translation="MAPKNCHGNDWKNQRSNLGENIVMAPYOSFHIGLGLAASVW*  
*PISACATPSEPPAVEQDQEVQLAQNSITYTEFPSPDEPKLNGKVEKG*  
*DYASISYRQEKKDCYSPPTVYDSESPMPALOTDMDRNPY*  
*PAVFAIYGGAHCCNEVKFTSNGSDSVRDGFNGPHPAPBDLNNGWDKEVE*  
*VDNRTPLFESYASPAJWALQWVDFVSFSFOHTHRAQSMENKDPEIWA*  
*QDNSSNGELAYKVALIGEDEWOMTLLKYYDRESDWLGLTCLYDDOSNCNLNEV*  
*YDSDFDALRFLVNEAGIYEAKFEAEMNT"*  
*complement(4'338..5'093)"*  
*/note="ORF\_ID:s110996"*  
*/codon\_start=1*  
*/product="hypothetical protein"*  
*/db\_xref="PID:91652230"*  
*/transl\_table=11*  
*/translation="MNVSHRYHITFGCOKNAKDSERMAIGILEMIGMTTDDPROAD*  
*LVMNMCSTDNAEOKYSLVGRKQKROPELWYACVYRQEGQQLRRVYED*  
*LVMGQPHANRUDQOLQEQWVAGSQWVATESHIMEDITKPRESTTSAWNNIYSGNER*  
*CSYCYTPNVRGVEQDERTPELYGEVLAQGKFETVLSQNTDAGYDRDGPSPSGR*  
*HLHILTDLXHVHDIGLEPRLTSHPELQOCBLKVCCEFFNIPPSGDN*  
*DILKAMKRGTRKLTQIIRKTRMPDAISADVTFQGETEAQFENTNLBEGV*  
*FDLUNTAASPRCPSPAAFWNOOSEEEVKEDRORLNUHSTOAMERSOYLGRVEEV*  
*LVESENLKSPGQWMGRTRGRNRLTFOGEISESELIGKTVPVKITERAFSLTGAEALSV*  
*A" 5818 .. 6969*  
*/gene="sqab"*  
*/note="ORF\_ID:s1r1020"*  
*/codon\_start=1*  
*/product="sulfolipid biosynthesis protein Sqab"*  
*/transl\_table=11*  
*/translation="MRAVIGGGCGWATAYLSNKGYEVGILDSLVRYWDALQ*  
*ETLUPAPIQRDLWYELJGKDKIDFQH*  
*PF5MIDREHALVLTANNVGNLNLTYALKDIFPCHLKVKGTMGXYGTNIDBEGYI*  
*TIEHGKCPDFLPIPKSYTHSKVHSNITFACKNGSRIDTQDNGVYGVITEE*  
*TGMDEMILNLHDYDVGFTGALNRFCQIAHGHPKGRCGQTRDYMCIEL*  
*AIANPADKGSOFVNEYOTLEFSVSDLAQOMYOKAGADGLKVEIDLENPVELBFHY*  
*NAVWNLDLGLQPLFLSDSLLSLLNFATKYKDVQKHILPKWTWG" 7112 .. 7402*  
*/note="ORF\_ID:ssr1698"*  
*/codon\_start=1*  
*/product="hypothetical protein"*  
*/db\_xref="PID:91652232"*  
*/transl\_table=11*  
*/translation="MADPITPAIDRICKHNEDHASATIYAQVFQGQDTVTHAQMQ*  
*ALIDTGMDLUVYESGGSKTRIERFQOPLKQDSEDAHQVLTAMAKARSVGKNS" 7439 .. 8134"*  
*/note="ORF\_ID:s110995"*  
*/codon\_start=1*  
*/product="hypothetical protein"*  
*/db\_xref="PID:91652233"*  
*/transl\_table=11*  
*/translation="MVALDREIYIPDSQPMADNTEOFEWIULKENIECLEFAHP*  
*DVFVGGDLWYPVAGHPELRVAPDMVAQGRPKRGSRQWQENQAQVFEFLSP*  
*GNTIKEMTRKLFDFHGHGTEYYYDPDNELLTQIGELTTEEMAHWSLIGI*  
*KFELSAETRVVYDGPRFLSTVALAQEQASORANEEAQARERKRAELAEND*  
*RKKAIAEAGDV" 15*  
*/note="ORF\_ID:s110994"*  
*/codon\_start=1*  
*/product="hypothetical protein"*  
*/db\_xref="PID:91652234"*  
*/transl\_table=11*  
*/translation="MPSINFVR1QHGYCOKTIVSINTVILPHLANLTWYLIG*  
*LGICLGMITSGDDDEVMEJLAEASSDHFVLDVQDQGDMEEGEAPMQLQ*  
*LSFFGIGKQPLMLIGDLSWLGIGWLNVTWVTPMREPLWAGWIFLSSLAT*  
*SLWGLRLASRPIALFKTSQDVTTERVIGCTGVTSKLPYLNGLTGOAHVVDNAG*  
*NLTIVSVPDWAVVPHANEQELIIDOSPSKGIGNLAIKSDDEDKWLS" complement(8'932 .. 1'0059)" 1..1344*  
**CDS**  
*/note="ORF\_ID:s110993"*  
*/codon\_start=1*  
*/product="Potassium channel"*  
*/db\_xref="PID:91652235"*  
*/transl\_table=11*  
*/translation="MGLOSSOENLNLLNLDQGRRLQELMAGATIAGLFLVUGTAWI*  
*RIVEDWNLWDLARYMTTITLTCGCGTSGQIAFEAVENIPFWVTDASPE*  
*AFOGFDSSRRROSKVTRLADWILCCTGFGTGFGEHPLSPASRIFTITLMLGTTGMYNRT*  
*VTLQALRDYVQLGDATDELAHTRATCIVSLSIDEVPLVSLAKLNKU*  
*RATRASSEAVOKLKRAGADEVVPSITGGRLLAARALRQVVSPLVSDGTDIFSP*  
*YMEERIGADECPTYGOTLARQLRQSGALILATRQDKLKVIGPMDTHLDAUSL*  
*ICLGFQQLRNLQICPLNPFRVLPKMR" 10183 .. 11472*  
*/gene="argD"*  
*/note="ORF\_ID:s1r1022"*  
*/codon\_start=1*  
*/product="N-acetylornithine aminotransferase"*  
*/db\_xref="PID:91652236"*  
*/transl\_table=11*  
*/translation="MYSPPWVSEVSEQAFTVDSLPAEFATPDTYVMMTYGEPPI*  
*ALARGOSTLNDTEGSKYSLDVGAGTCIGHAVERVLSQDFOKLHRSNLNTYP*  
*EOGELAKWVTHSCADRFVCEGDAANEKATKLVYAHVFLDPEQVLTAKSF*  
*HGRLLATITTGQPKYQOYQFPLVPGDVFYVNDYRSLENKVADEGENSWAATF*  
*PLQGEQVREPDGLATKVRKSCQNDLIVDFEVYVGRTGKJYHVEGDPD*  
*TSAGKLAGGVITGAMCKKCFDVFEGNHASTFGGFLPLACGAVLTKTLEGDPD*  
*VOAQRGQLRSLAEITKNOYPLTEVGRWGLINGLEISASSLTSVEIVRAAMEQGL*  
*LAPAGRKVRLVPPVLTERRAIAQVEILRQIAITLV" 11599 .. 11913*  
*/note="ORF\_ID:s1r1023"*  
*/codon\_start=1*  
*/product="hypothetical protein"*  
*/db\_xref="PID:c1652237"*  
*/transl\_table=11*  
*/translation="MENLWQDPLDQDQEOPDLENNKNITLERIISTGQITPPGEWY*  
*DODPREWVWVILQKABITYTGAVISLQGDYLLIPARQHRVTFTHEPCIWALH*  
*FS" 11920 .. 12161*  
**CDS**  

Note: remainder of annotations omitted.

Query Match						
Best Local Similarity		Score		DB		Length
81.6%		24		20		150894
Pred. No.		1.38e+00		0		Mismatches
0;		0;		0;		Indels
Conservative		31;		0;		Gaps
Matches						0;

Db 4235 aaacctggatggaccatcaatccatccccaa 4272  
 Qy 1388 AACATCTAGATGGAGAACACTAACCTCCCAA 1425

RESULT LOCUS ECACEA 1344 bp DNA BCT 12-SEP-1993  
 DEFINITION E. coli aceA gene coding for isocitrate lyase (EC 4.1.3.1).  
 ACCESSION X0743  
 NID 940884  
 KEYWORDS aceA gene; isocitrate lyase; isocitratase.  
 SOURCE Escherichia coli.  
 ORGANISM Escherichia coli.  
 Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 KEGG-SEA  
 REFERENCE Cozzzone, A.J.  
 AUTHORS Cozzzone, A.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAY-1988) Cozzzone, A.J., University of Lyon,  
 Laboratory of Molecular Biology, 43 Boulevard du onze Novembre,  
 6962 Villeurbanne, France  
 2 (bases 1 to 1344)  
 REFERENCE Cozzzone, A.J.  
 AUTHORS Cozzzone, A.J.  
 TITLE Nucleotide sequence of the aceA gene coding for isocitrate lyase in  
 Escherichia coli.  
 JOURNAL Nucleic Acids Res.  
 MEDLINE 16 (12), 5689 (1988)  
 FEATURES source  
 source 1..1344 Location/Qualifiers

```

<organism="Escherichia coli"
/stain="K12"
/map="91 min"
27 .30
/note="pot. rRNA binding site"
40 .1344
/note="isocinrate lyase (AA 1-434)"
/codon_start=1
/db_xref="PID:940885"
/db_xref="SWISS-PROT:PO5313"


```

GIN COUNT 321 a 353 c 386 g 284 t

Query Match	1.5%	Score	23	DB	20;	Length	1344;
Best Local Similarity	81.1%	Pred. No.	5.59e+00				
Matches	30;	Conservative	0;	Mismatches	7;	Indels	0;
						Gaps	0;
Db	983	cttataactgtcccgtcgttcaactggcagaaaa	1019				
Qy	1488	CTTATTACTGCTGAGTAATTCACTGGAAAAAAA	1524				

Search completed: Thu Aug 21 10:01:07 1997  
Job time : 2597 secs.

**THIS PAGE BLANK (USPTO)**